K. K. Wagh Institute of Engineering Education & Research, Nashik

(An Autonomous Institute From A.Y. 2022-23)

WINTER-2025	
Exam Seat No.:	
Academic Year:2025-2026	Semester:V
Class:TY	Program:B.Tech
Branch Code:INT	Pattern:2023
Name of Course:Machine Learning	Course Code:2308303
Max. Marks:60	Duration:2.30 Hrs.

Instructions: Candidates should read carefully the instructions printed on the Question Paper and on the cover page of the Answer Book, which is provided for their use.

1. This question paper contains _____page(s).
2. Answer to each new question is to be started on a new page.
3. Assume suitable data wherever required, but justify it.
4. Draw the neat labelled diagrams, wherever necessary.
5. The last columns indicates the Course Outcome and level of Blooms Taxonomy of the Question/sub-question.

Marks CO

Question No. 1

a) Describe the different types of scales of measurement and illustrate each with an appropriate example. (6) CO1

Question No. 2

b) Analyze the model for computing Accuracy, TPR, and TNR, and explain how these metrics help evaluate a binary classifier's performance. (6) CO2

Question No. 3

a) Explain the regression coefficient and discuss SSE, MSE, and MAE in the context of regression. (5) CO3

OR

b) Identify the attributes of an optimal regression line and discuss their significance. (5) CO3

c) Describe underfitting and overfitting in regression models, highlighting their implications. (5) CO3

OR

d) Differentiate between ordinary least squares (OLS) regression and regularized regression. (5) CO3

e) A researcher wants to study the relationship between the number of hours students studied (X) and their exam score (Y). The following table contains the data for 7 students. Compute R^2 error of the best fit line "Y = –1.8X + 46". (6) CO3

Sr. No	X (Hours Studied)	Y (Exam Score)
1	2	42
2	5	38
3	7	33
4	9	29
5	11	26
6	13	20
7	15	18

OR

f) Consider the following dataset of x and y values: (6) CO3

i) Determine the regression coefficients β_0 and β_1 for the simple linear regression model that best fits the given data.
ii) Write the regression equation and interpret its meaning.
iii) Using the regression model, estimate the value of y when x = 90.

	x	y
1	75	85
2	85	95
3	70	80
4	80	65
5	70	75

Question No. 4

a) Explain the following measures of impurity with example (5) CO4

i) Information Gain ii) Gini Index iii) Entropy.

OR

b) Illustate the working principles of XGBoost and analyze how its sequential tree-building process enhances model accuracy. (5) CO4

c) Discuss the concept of decision trees and illustrate the application of the C4.5 algorithm. (5) CO4

OR

d) Analyze the working mechanism of the Random Forest algorithm and evaluate how its ensemble structure contributes to reducing overfitting in predictive models. (5) CO4

e) For the given data set apply Naïve Bayes Classifier and predict the Class for weather = Rainy and car = Stay-home. (6) CO4

Weather	Car	Class
Sunny	Working	Go-out
Rainy	Broken	Go-out
Sunny	Working	Go-out
Sunny	Working	Go-out
Sunny	Working	Go-out
Rainy	Broken	Stay-home
Rainy	Broken	Stay-home
Sunny	Working	Stay-home
Sunny	Broken	Stay-home
Rainy	Broken	Stay-home

OR

f) A company wants to predict whether an employee will take Work-from-Home (WFH) based on three factors: Rainy Weather, Traffic Jam , Health Issue. The HR team collected the following dataset from 10 employees. (6) CO4

(a) Compute the entropy of the target variable WFH

(b) Calculate the Information Gain for attribute: Traffic Jam, Health Issue.

Question No. 5

a) Apply KNN on the following data and classify the new sample (3,5) to the respective class. What will be the effect on output if k = 3? (8) CO5

X	Y	Class
7	7	Pass
7	4	Pass
3	4	Fail
1	4	Fail
4	3	Fail
6	7	Pass

OR

b) Cluster the following eight points (with (x, y) representing locations) into three clusters: A1 (2, 11), A2(3, 5), A3(8, 5), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9). Use Euclidean distance as the distance metric. If the initial medoids are A1, A3, and A7, determine the cluster medoids after the second iteration. (8) CO5

c)	Interpret the terms Rule, Support, Confidence, and Lift in the context of association rule mining, and illustrate how each metric influences decision-making using suitable examples.	(8)	CO5

OR

d)	Describe the concepts of : 1) Euclidean Distance 2) Manhattan Distance 3) Hamming Distance 4) Minkowski Distance Metric with suitable examples.	(8)	CO5

····· End of question paper·····